User Manual

Author: CHEN LIU (C00140374)

Supervisor: Paul Barry

Date: 16th April 2010

Content

1. System Requirement

Anit-SpamMetric system runs in python2.6 pre-installation environment (requires Windows2000/xp/2003/2008/Vista, MS-DOS). You can download python2.6 from website http://www.python.org/download/. It is free to download and install. You can install it anywhere in your system. Email server is required. You can install a local host email server such as CMailServer, or use an external email server. The SMTP/ESMTP and POP/IMAP server is supported in this system. If installed firewall, please make sure open the transport port that depends on what you set in your email server.

2. Interface and Port Configuration

Before using the system, please research some information at first. We will tell you how to fill a correct SMTP and POP information in this system.

i.   Transfer Server
      Server interface, port and the type of server (SMTP, ESMTP)
ii.   Receiving Email Server
      Server interface, port and the type of server (POP, APOP, IMAP)
      This system only supports POP and IMAP email box.
iii. Anti-Spam filter or Classifier
      This system can give a simple estimation to the type of anti-spam filter (classifier) that is running at that time. It could guide user to configure the information set.
      Filtering Type:
      SMTP filter will run when system try to connect the specific SMTP port and prevent spam before email arrive in mail box. Please set the same incoming interface and port of SMTP filter as this system, and outgoing interface and port will be the same as your email server.
      An example about how to configure SMTP filter in CleanMail and localhost CMailServer

In this step you have to specify the interface and port number where CleanMail will listen for incoming SMTP connections, and the address and port where your mail server (the outgoing server) is listening.
Choose '<all interfaces>' if you want CleanMail to listen on all interfaces. The default SMTP port number is 25.

Incoming IP   <all interfaces>   Any Free Port

Incoming Port   25

CleanMail forwards commands and messages to the outgoing SMTP server (your mail server).

                                                           Local Mail Server

Outgoing Server   localhost

Outgoing Port   30   Test

Advanced

Internet Mail | Account | Log | Email Filter | Miscellaneous

Language Selection   English
When Windows Startup
  Minimized   Autostart   Autolock
SMTP Port   30
POP3 Port   110
Account Port   8011

Anti-SpamMetric System SMTP port option

Information Setting

Please set sender, receiver, smtp and port.

                                                    CMailServer
                                                    Domain

Sender: admin@sara-01876ef798.com

Password: SMTP transport

Receiver: admin@sara-01876ef798.com   Reciever
                                                    Address
SMTP: localhost

Port: 25

         OK   Cancel

Pop filter will run when this system checking the same interface and port. You should configure the same information in both pop filter and this system.

Pop to SMTP filter will run when you tell the filter running. It could run check manually or on filter own. This system could not change the time of checking but user can. You should tell filter your email account information and checking address and port, filter will retrieve email from your email box. You should set the same SMTP information as your email server thus filter will deliver legitimate email back to you.

Classifier:

If you use an anti-spam classifier rather than a filter, in other words, the classifier could not isolate spam from your email server you should set or check how the classifier to mark spam email. And add a new checking rule to this system. This system could check the email classification of anti-spam classifier as well.

3.Zip Folder Requirement

a) Files are all TXT files in the folder.

The subject of email is as TXT file name. The content of email is as the content in txt file. Different email is in different txt file. This system doesn't support attachment and image email.

b) Different type of email in different folder

Please put different type of email (spam or ham) in different folder, if you want to correctly evaluate anti-spam software.

SPAM Type:

HAM Type:

c) Email files compress to Zip Folder

All files need to be packed in a zip folder (*.zip).   Zip folder name is not restricted, it can be any names. But you should remember which type email in the folder.

4.How to run

This system is simple to be operated. We assume that python2.6 has been installed in the path G:\python2.6, you should put the entire folder into the path. Then run in accordance with the installation

1. Open CMD command line
2. Enter the python directory.

    `G:\python2.6>`

3. Enter this system folder directory

    ```
    G:\python2.6>cd Anti-SpamMetric System

    G:\python2.6\Anti-SpamMetric System>
    ```

4. Run AntiSpamMetric.py
    Run antispam.py file in the directory of file folder.

    `G:\python2.6\Anti-SpamMetric System>AntiSpamMetric.py`

    Sometimes if you have installed python2.6, you could skip these steps from 1 to 4 and run it by double click AntiSpamMetric.py in this system folder. This is not recommended. Some unknown error might occur.

5. Configure SMTP Option(Details see SMTP configure section)
    Click Send->SMTP Option

    If sending email uses SMTP, Password column is filled empty.

If sending email uses ESMTP, Password column should be filled. The sender address will be different with receiver address.

Otherwise, a socket error will happen.

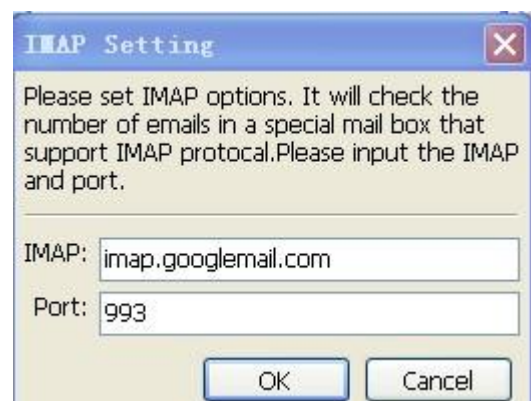6. Do not worry to send email. Configure POP /IMAP Option(Details see the section of POP configure)

7. Clean email box using POP/IMAP.
    Click Check->CleanMail
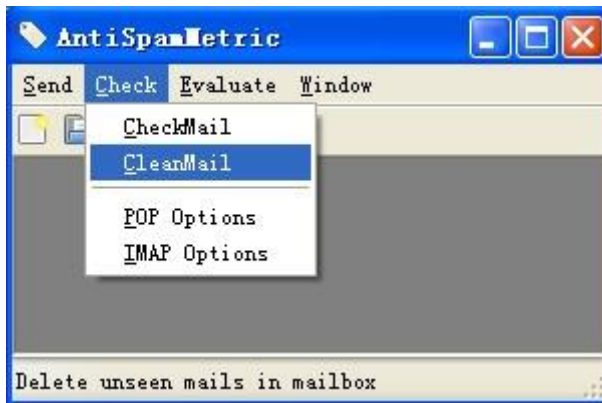
Email address of user will be the same as receivers' in SMTP setting. Don't forget to choose mailbox type.
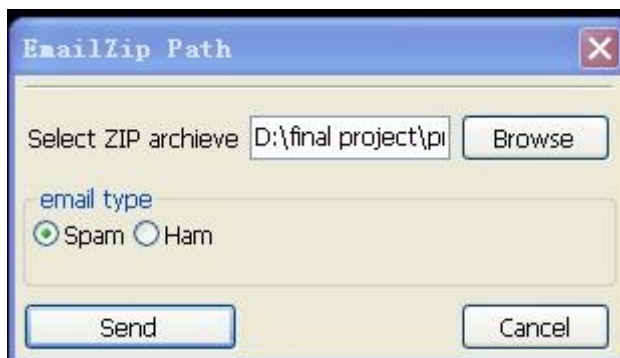

8. Send ZIP folder

Before sending email, make sure your anti-spam product is running and SMTP information is correct. Click Send->SendMail.

Upload a ZIP folder. The requirement of ZIP folder is in the 3 section, see folder requirement.

Choose file folder type.

Click Send button. Sending progress.

There is a shortcut key for sending Email in the first place of toolbar on the left.
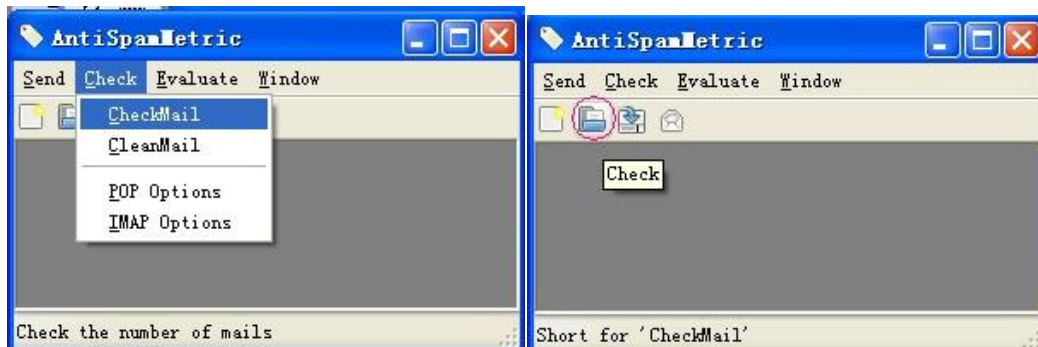
Send Succeed

9. Check using POP/IMAP.
   Click Check->CheckMail or shortcut at the second place in toolbar on the left for checking email box.

Email address of user will be the same as receivers' in SMTP setting. Don't forget to choose mailbox type.
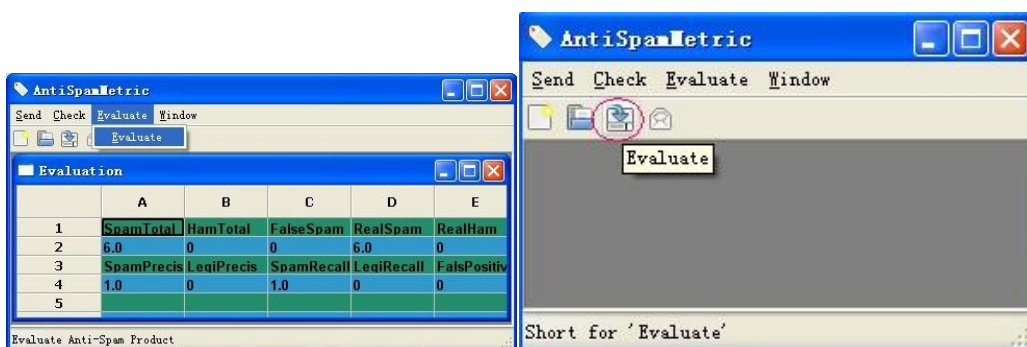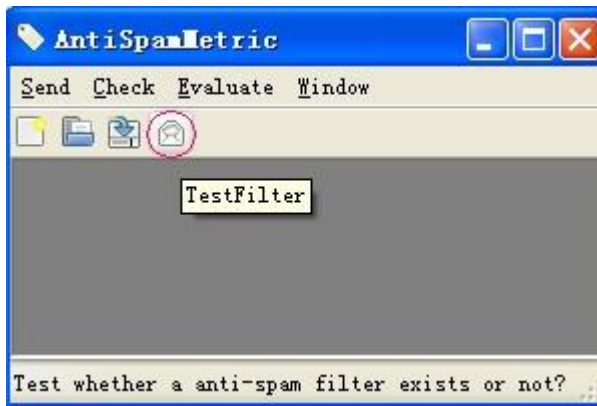
Check Succeed:

10. Evaluation

Click Evaluate->Evaluate or click shortcut for evaluate at the third place of tool bar on the left.

11. This system could give an estimate for filter.
Click the last shortcut button on the left.

As a result, you should make sure there is no unseen email in your email box otherwise press the clean mail button under pull-down menu of checking email. And make sure port and interface information is filled correctly remember to open an anti-spam product that you want to evaluate. Then, upload two different types of email zip file separately and check email box. At last, press the button of evaluation, the results will display.

5. User defined Rule

It is an extra function. If you use a classifier rather than a filter such as SpamBayes you can use this function to define a rule to implement email classification. If you do not use a classifier, you could ignore the Rule's form, leave it empty. It will make no useful to delete and test function, so when you run delete or test function you could ignore this RuleSpam and RuleHam form.

When you run a checking function, and want to use a classifier. Fill in the field form as follow.

Make sure the key word is correct, otherwise please leave it empty.

Usually, we need not to use an email classifier. Just leave it as follow.

6. Evaluation

Spam Total is the amount of junk emails that are sent at last time.

Ham Total is the amount of normal emails that are sent at the last time.

False Spam is the amount of spam that are incorrectly classified in inbox.

Real Spam is the amount of spam that are correctly classified in junk folder.

Real Ham is the amount of normal emails that are correctly classified in inbox.

False Ham is the amount of normal emails that are incorrectly classified in junk folder.

The performance evaluation:

1.   Spam Precision(SP)

Spam precision is the proportion of mail messages classified as spam that real spam. The more precision is, the better performance of filtering spam is possessed.

2.   Legitimate Precision(LP)

Legitimate precision is the proportion of mail messages classified as legitimate that wanted mails (ham). It is an opposite side of false positive. The more this precision is, the better capability of discovering legitimate has.

3.   Spam Recall(SR)

Spam recall is the proportion of spam mails that are classified as spam. It is a correct precision in the process of filtering and classifying spam email.

4.   Legitimate Recall(LR)

Legitimate recall is the proportion of good mails correctly classified as legitimate mails. It is a correct precision in the process of analyzing and filtering legitimate email.

5.   False Positive (FP)

False positive is the proportion of legitimate mails incorrectly classified as spam. It is a precision of false legitimate email classification, which is dangerous if the value is too high. It might filter an important email that you except strongly to check.

6.   False Negative (FN)

False negative is the proportion of spam incorrectly classified as legitimate mails. It is a precision of false spam classification. It could be allowed to happen since a filter could never be smart as people. If the less precision it is the less spam you will receive.

Both false negative and legitimate recall are two well known measures used in anti-spam classification. They are main measure metric of anti-spam technologies. According to the result, we will obtain the comprehensive performance evaluation of anti-spam technologies.

Screenshot

It is a screenshot of evaluation to McAfee SpamKiller product.

	A	B	C	D	E	F
1	SpamTotal	HamTotal	FalseSpam	RealSpam	RealHam	FalseHam
2	200	230	14	186	144	86
3	SpamPrecis	LegiPrecis	SpamRecall	LegiRecall	FalsPositive	FalsNegativ
4	0.683823529412	0.911392405063	0.93	0	0.373913043478	0.07

7. Trouble Reservation

If Anti-SpamMetric system appears a no response state in a short time while you send lots of emails, or checking emails, do not be anxious to kill the process. Give it a minute or two to allow the system finishing sending email or checking the mail server.

8. Some Common Email Server Configure

There is a little time delay on Gmail server. Some email boxes need TLS encryption. This system only supports gmail and hotmail.

Please make sure sender address is different from receiver address in 163 mail server.

Test Mail Box:

Mail Server	Interface	Port
Gmail	smtp.googlemail.com	587
Gmail	imap.googlemail.com	993
163	smtp.163.com	25
163	pop.163.com	110
163	imap.163.com	993
qq	smtp.qq.com	25
qq	pop.qq.com	110
hotmail	smtp.live.com	587